

# Your Second RecSys

**Сафило Ильдар**

Руководитель группы рекомендательных систем  
МТС BigData



# Бизнес-эффект от рекомендаций

# План

Как рекомендации влияют на продукт?

А/Б тест

Связь офлайн и онлайн метрик

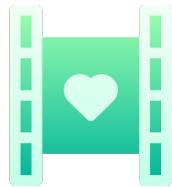
# План

**Как рекомендации влияют на продукт?**

А/Б тест

Связь офлайн и онлайн метрик

# Какие метрики оценивать в рекомендательных системах?



Фильмы



Банки



Социальные  
сети



Поиск  
работы



Книги



Клики  
Просмотры  
Лайки  
Покупки



Оформление  
заявок  
Приобретения  
банковских  
продуктов



Добавление в  
друзья  
Время в  
приложении

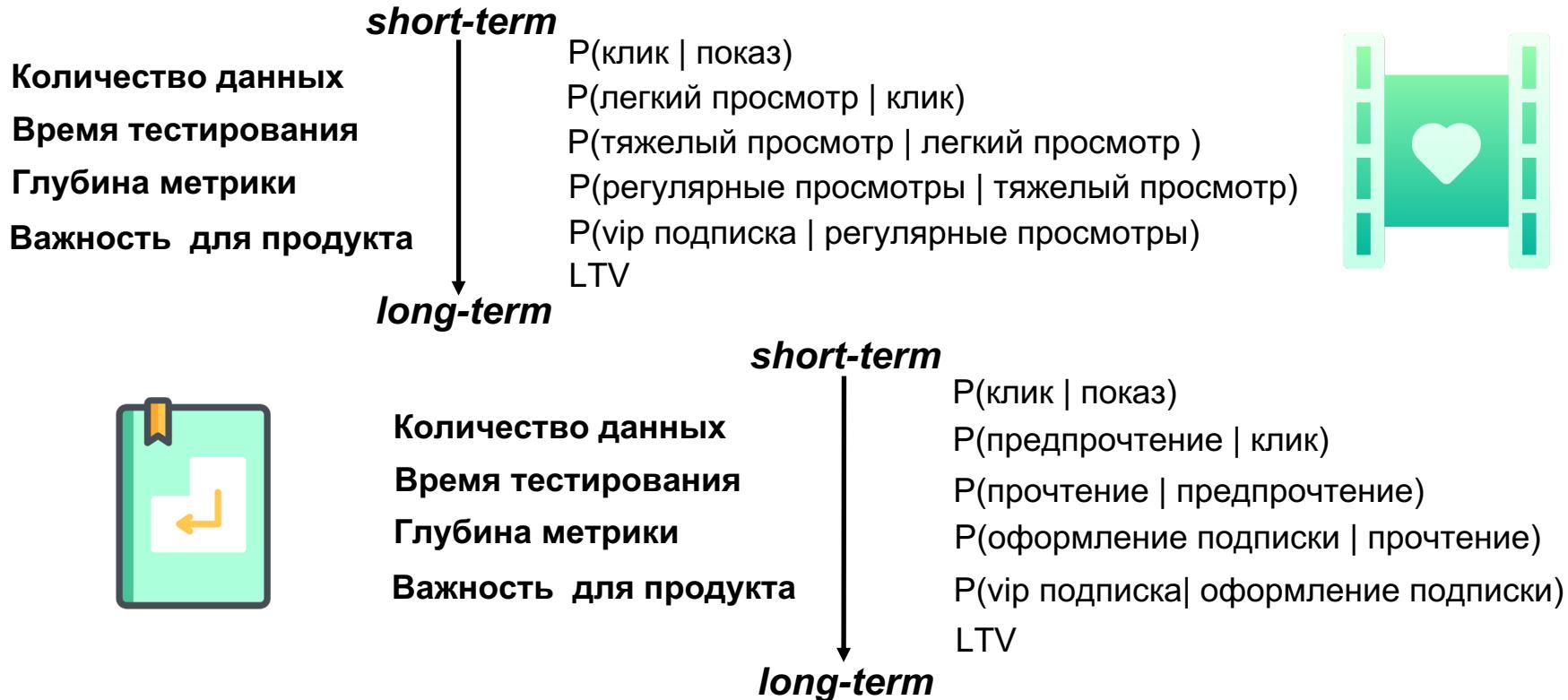


Клики  
Созвоны с  
рекрутерами

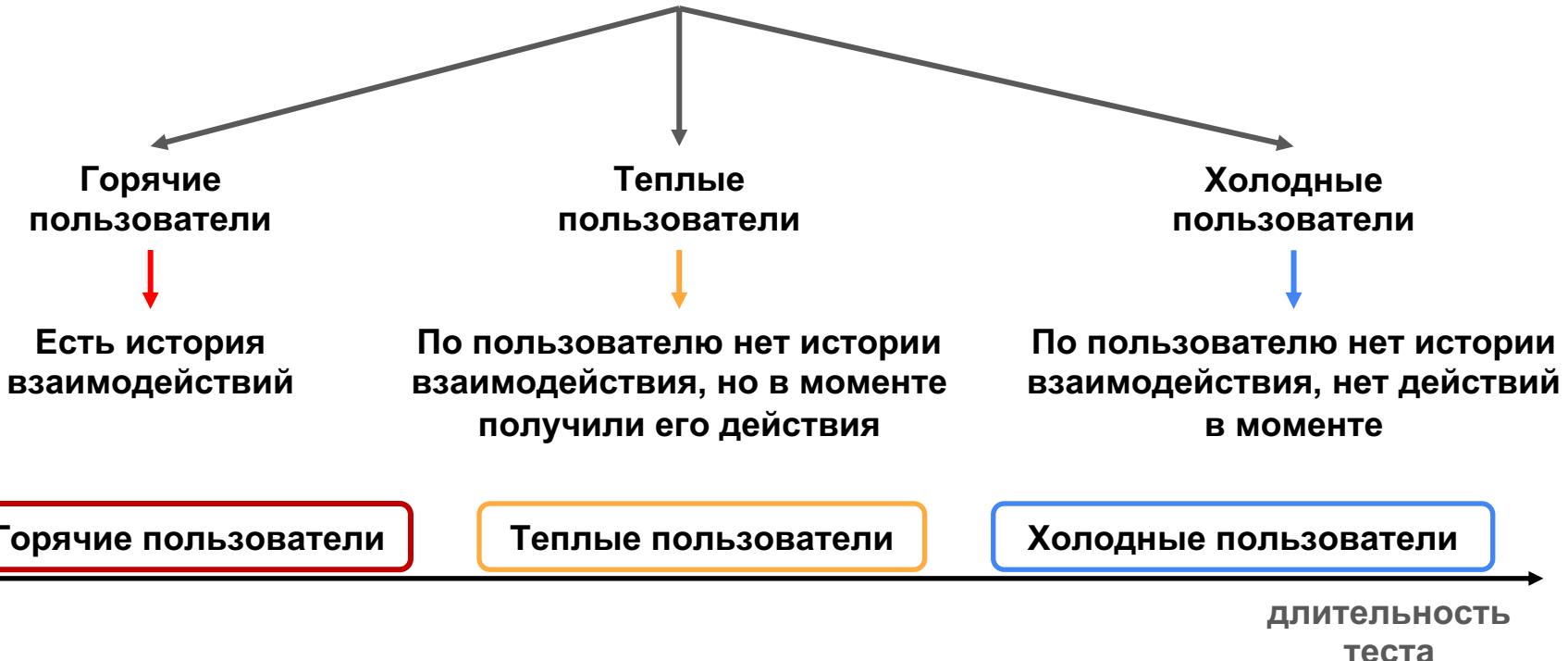


Клики  
Прочтения  
Переход в  
аудио книги

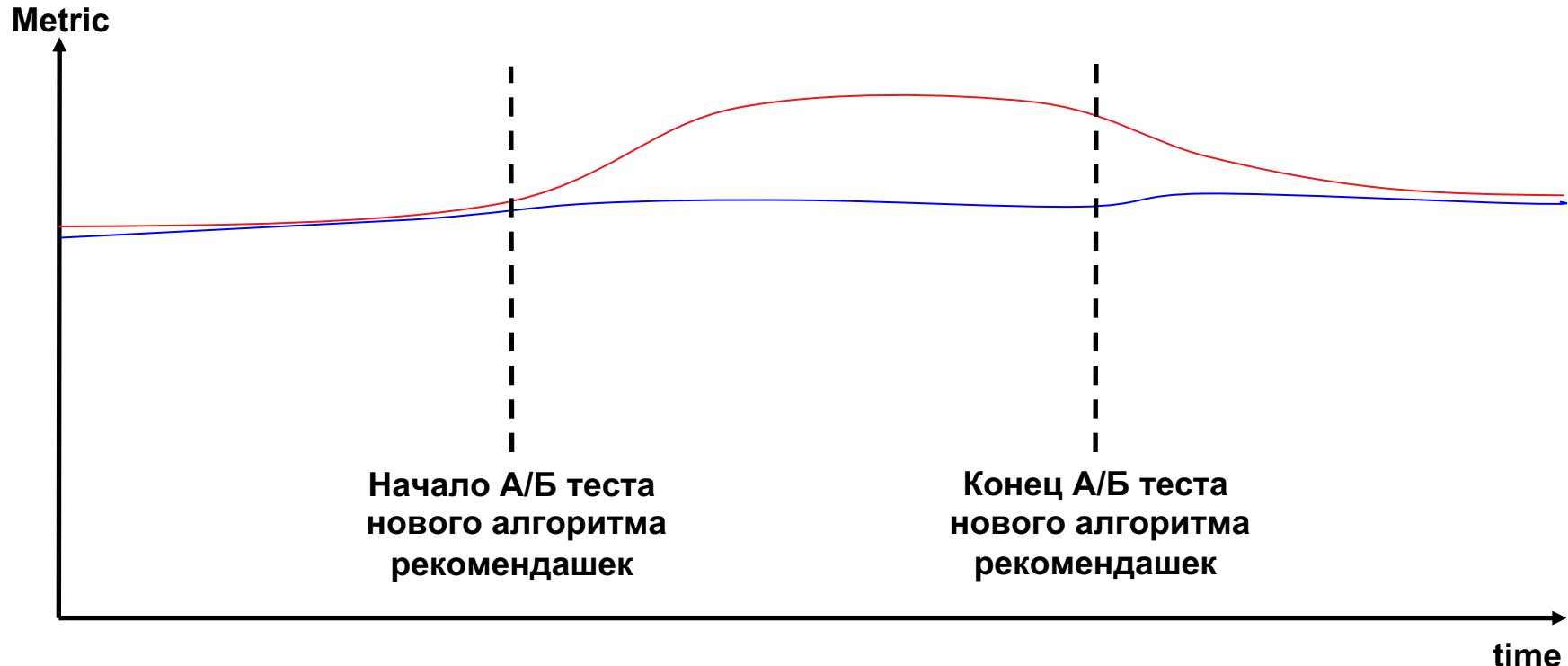
# Быстрые / долгие метрики



# Разные группы пользователей



# Эффект новизны



# Где размещать рекомендации?

Размещение блока с  
рекомендациями может быть более  
важной  
задачей, чем сами рекомендации!

## На главной странице пользователей

- специальная полка рекомендуемое
- персонализация контента внутри  
других полок новое / популярное /  
в тренде

## Рекомендации похожих айтемов/товаров на странице айтема/товара

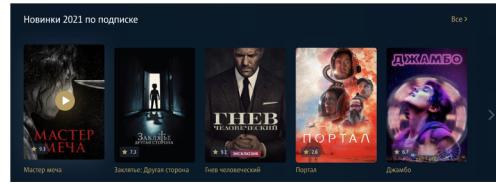
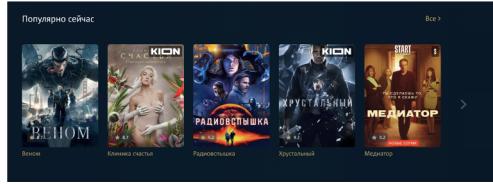
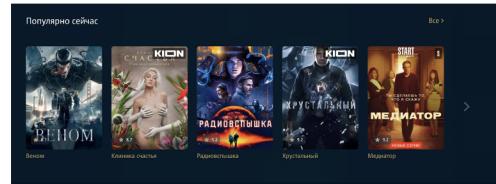
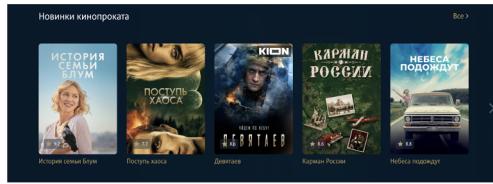
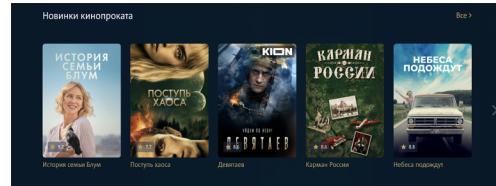
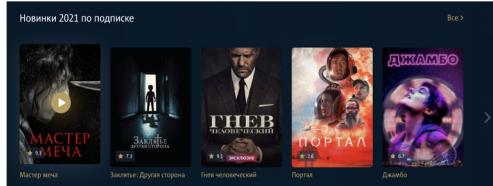
The image displays three distinct mobile application interfaces, each illustrating a different strategy for integrating recommendations:

- Top Screen (Recommendations):** Shows a navigation bar with categories like "Сыщик" (Investigator), "Классик" (Classic), "Странник" (Traveler), and "«Большой брат»" (Big Brother). Below is a grid of book covers with titles such as "Безмозглый пациент" by Алекс Михаилides, "Дэн Браун" by Дэн Браун, and "Приключения" by Роберт Глебройт. A call-to-action at the top says "Выберите категорию и мы подберем книги специально для вас" (Select a category and we will select books specifically for you).
- Middle Screen (Popular Now):** Features a dark header "Популярно сейчас" (Popular now). It shows a grid of movie posters with titles like "Веном" (Venom), "Клиника счастья" (The Clinic of Happiness), "Радиовспышка" (Radioactive), "Хрустальный" (Crystal), and "Медиатор" (Mediator). Each poster includes a rating (e.g., 8.7, 9.2) and a "View more" button.
- Bottom Screen (Similar Items):** Shows a header "С этой книгой читают" (Readers also like). It displays a grid of book covers with titles including "Луиза Ленни Королевство слепых" (Louise Penny The Kingdom of the Blind), "Нож Ю Несбё" (John Nesbø The Knife), "The One. Единственный" (John Marrs The One. The Only One), "Почти счастливые..." (Maria Metlytska Almost Happy...), "Камень из Ватикана" (Tatyana Ustynova The Stone from the Vatican), "Ю Несбё Королевство" (John Nesbø The Kingdom), and "Маринина Безупречная репутация" (Alexandra Marinina The Perfect Reputation). Each book entry includes a star rating and the author's name.

# Где размещать рекомендации?

**Размещение блока с  
рекомендациями может быть более  
важной  
задачей, чем сами рекомендации!**

- Баннеры / реклама
- Можно персонализировать главную страницу автоматически



## Рекомендуемые продукты



### Дебетовая карта Tinkoff Black

Зарабатывайте деньги, а не бонусы. Закажите карту с кэшбэком до 30% и ежемесячным процентом на остаток

[Формировать карту](#)



### Кредит наличными

До 15 млн рублей на следующий день. От 5,9% годовых

[Получить](#)



### Страхование КАСКО

Доставка за день. Ремонт стекол, фар и зеркал без справок

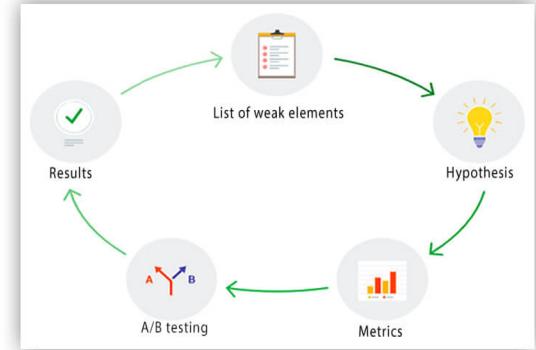
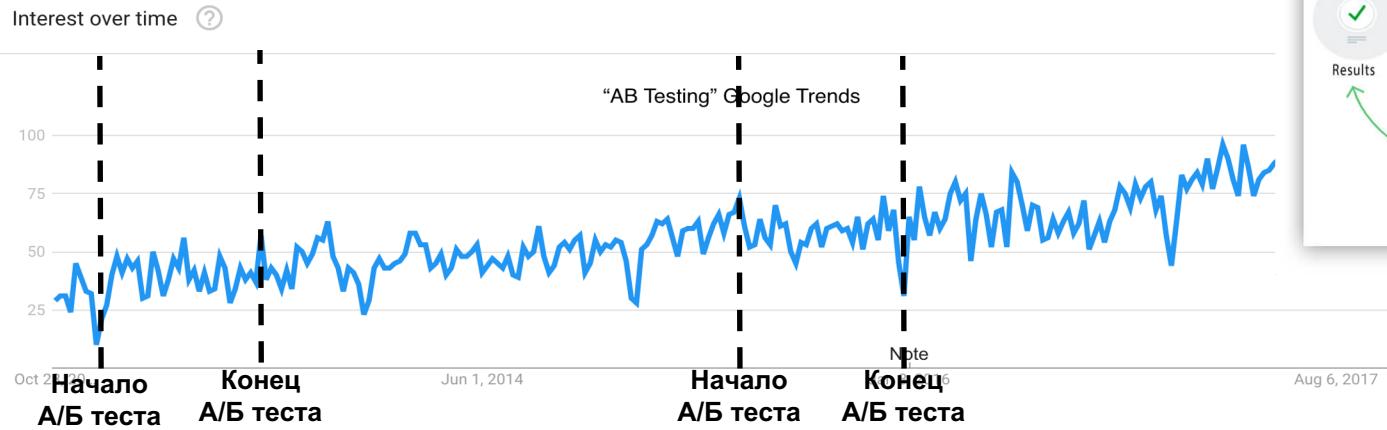
[Рассчитать стоимость](#)

Как рекомендации влияют на продукт?

**А/Б тест**

Связь офлайн и онлайн метрик

# Почему нельзя тестиовать сразу на всей базе?



# Зачем делать А/Б тест?

Количественная оценка влияния нашего алгоритма рекомендаций на продукт.

Один из “честных” способов доказать увеличение бизнес-метрик вследствие сделанных нами изменений.

Честная проверка любой продуктовой гипотезы.

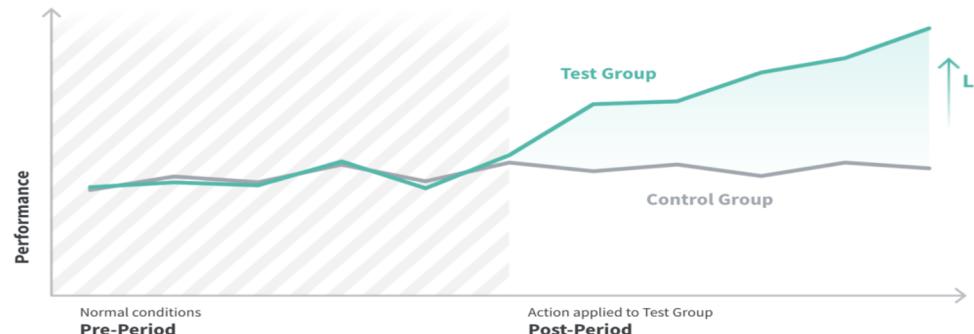
Статистический метод принятия решений.

Пример: Увеличились ли прочтения пользователей на 5% в среднем?

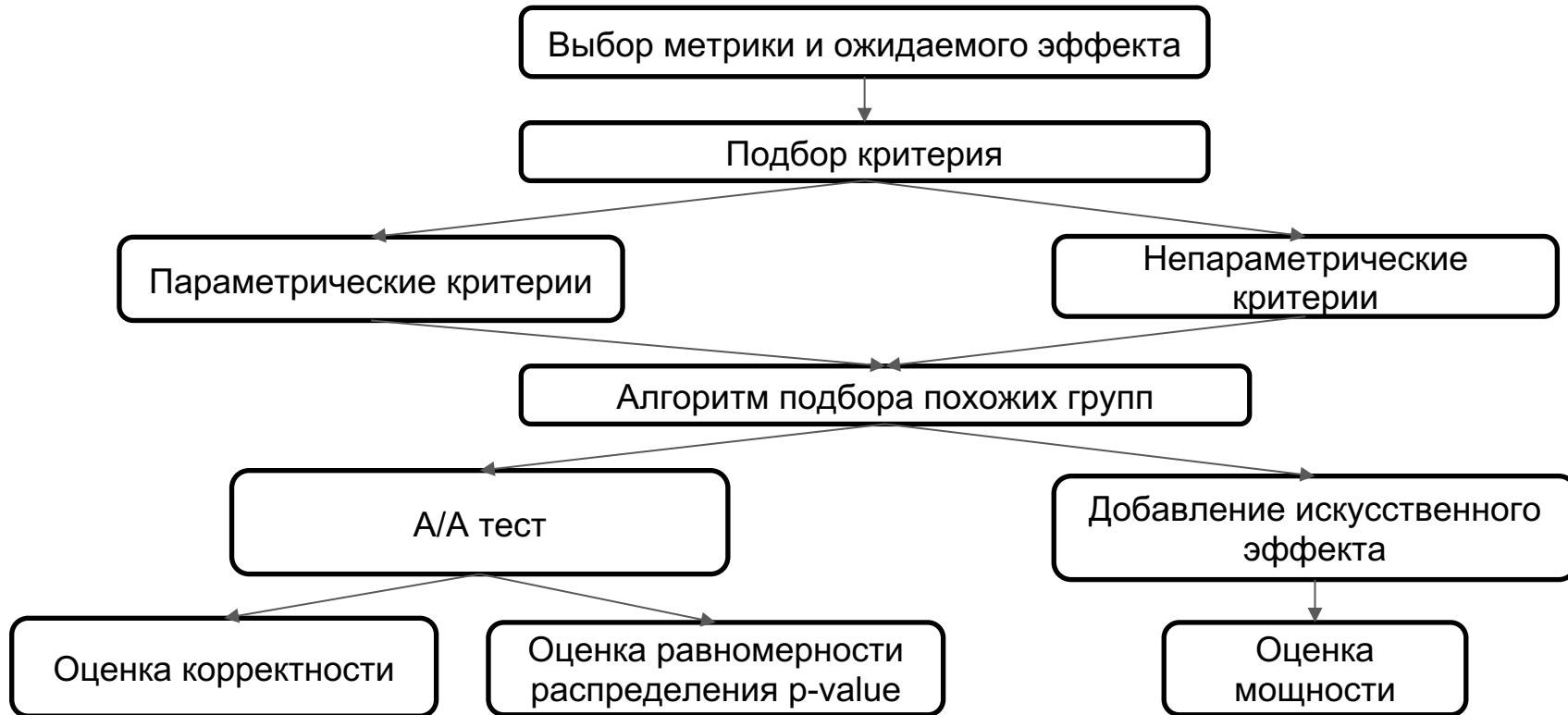
Пример: Раньше рекомендовали всем популярное, изменили алгоритм на ALS. Стало лучше или хуже?

Пример: Раньше рекомендовали всем популярное, изменили алгоритм на ALS. Стало лучше или хуже?

Пример: мы увидели, увидели увеличение CTR на 0.01% относительно случайного алгоритма, можно ли использовать наш алгоритм?



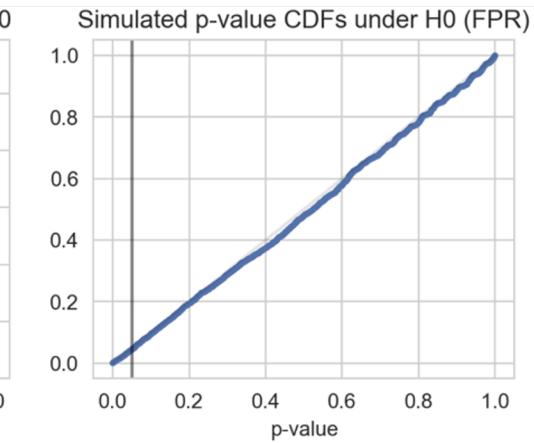
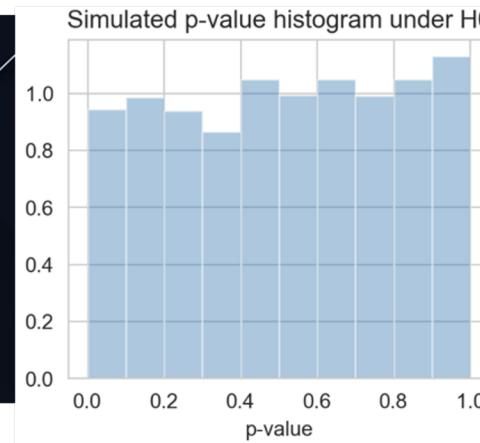
# Общий пайплайн А/Б теста?



# A/A тест как оценка корректности

Считаем процент пар, в которых  $p\text{-value} \leq \alpha$ : т.е считаем в скольких случаях критерий (сработал) нашел различия, когда его на самом деле нет.

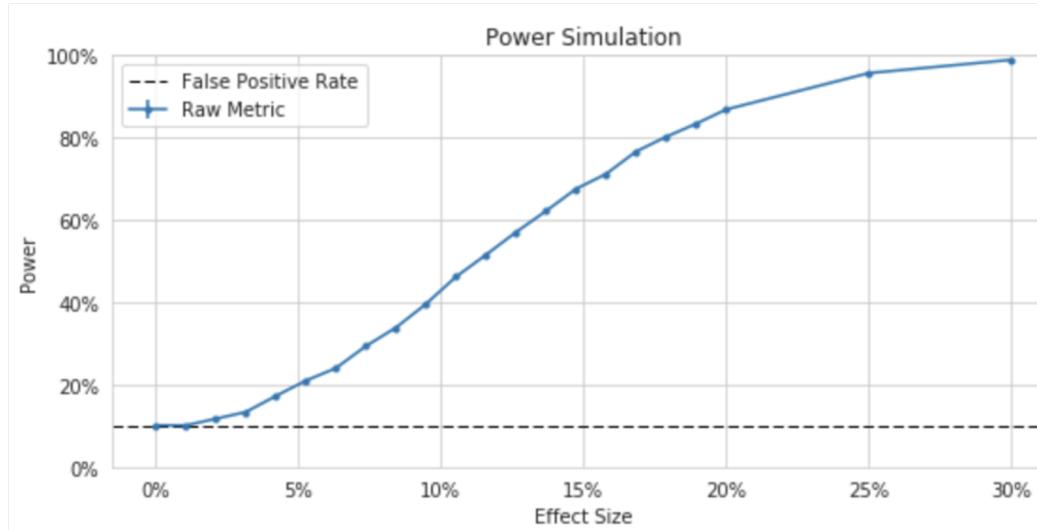
Процент пар не должен превышать  $\alpha$  для корректности нашего критерия и распределение  $p\text{-value}$ , которое мы получили после A/A теста на исторических данных, должно быть распределено равномерно. Без этого наши выводы будут смещены, критерий и A/B тест будут некорректными.



# Оценка мощности

Добавляем разные эффекты размером  $x\%$  в каждую пару в одну из групп

Считаем процент пар, в которых  $p\text{-value} \leq \alpha$ : т.е считаем в скольких случаях критерий сработал или нашел различия, когда оно на самом деле есть.



Как рекомендации влияют на продукт?

А/Б тест

**Связь офлайн и онлайн метрик**

# Исследователи в бою

Смотрим на одну любимую метрику (ROC\_AUC, MAP@k, precision@k, ndcg, ...)

- Оффлайн метрика не упала, но модель "нравится" -> надеемся на увеличение метрики
- Оффлайн метрика выросла -> наверное, должна вырасти и онлайн метрика?

Смотрим на несколько / комбинацию оффлайн метрик

- Какие-то оффлайн метрики не упали, а какие-то оффлайн метрики выросли, а модель очень нравится -> наверное, должна вырасти и онлайн метрика?
- Оффлайн метрики выросли -> наверное, должна вырасти хотя бы одна онлайн метрика?
- Оффлайн метрики не упали, модель нравится -> катим в тест, надеемся на чудо

Основная проблема в рекомендашках в отличие от классики – оценка на прокси метриках, которые не моделируют поведение пользователя

# Осторожные исследователи

Пытаемся найти корреляции / построить модель, которая бы по оффлайн метрикам предсказывала бы онлайн метрики.

Самое сложное – собрать датасет!

Какие действие могут помочь?

- Логирование оффлайн метрик и онлайн метрик на А/Б
- Устойчивая инфраструктура выкатки в прод
- Единый пайплайн обучения и валидации разных моделей

# Осторожные исследователи

**Единая группа пользователей разбивается на n частей 1 тест**

Маленькая подгруппа\_1 → ranker\_1 -> offline metrics\_1 -> online perf\_1

Маленькая подгруппа\_2 → ranker\_2 -> offline metrics\_2 -> online perf\_2

...

Маленькая подгруппа\_n → ranker\_n -> offline metrics\_n -> online perf\_n

# Осторожные исследователи

**Для short-metric делаем несколько А/Б тестов и  
оцениваем качество каждый день – считаем оффлайн  
и онлайн метрики**

алгоритм1 → offline A day1 -> online A day1

алгоритм2 → offline Б day1 -> online Б day1

алгоритм1 → offline A day2 -> online A day2.

- А/Б тест1 ... А/Б тест2 ...А/Б

тест n

алгоритм2 → offline Б day2 -> online Б day2

алгоритм1 → offline A day3 -> online A day3

алгоритм2 → offline Б day3 -> online Б day3

**Много А/Б**

A/B 1:

Offline A -> Online A

Offline Б -> Online Б

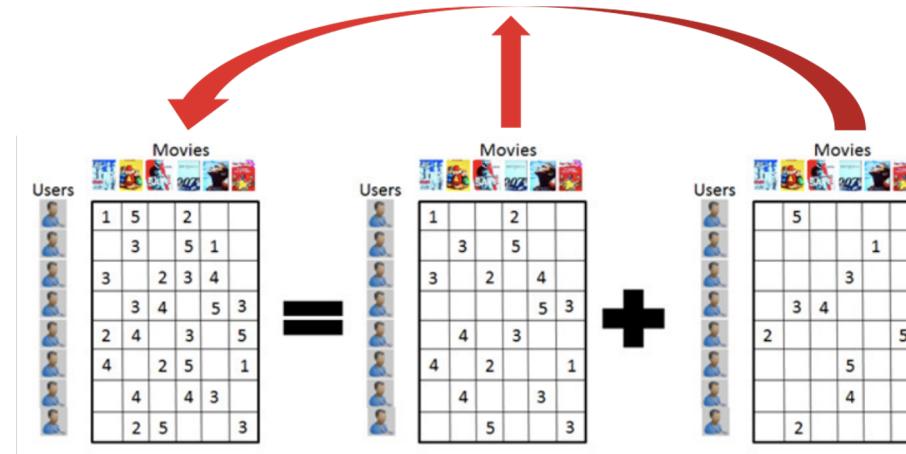
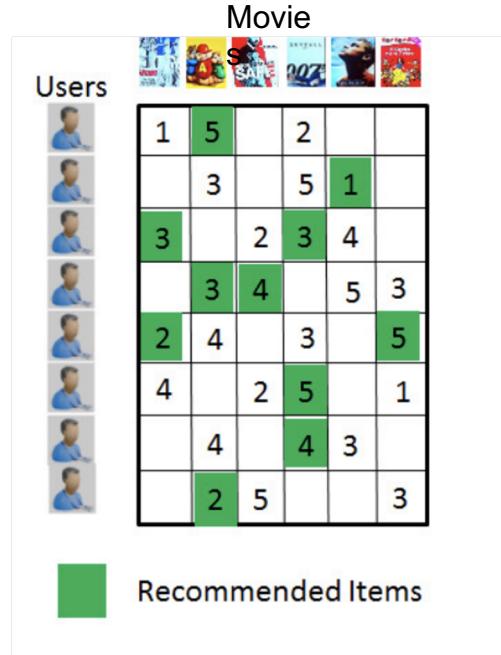
A/B 2:

Offline A -> Online A

Offline Б -> Online Б

# Bias в оффлайне

## Feedback Loop



# Feedback loop – решения



## Простые популярные решения

- Случайно подмешивать рекомендации из популярного/эвристики
- Тестиовать бандитов
- Глобальная контрольная группа без экспериментов
- Эвристики

## Сложные решения

- <https://papers.nips.cc/paper/2016/file/962e56a8a0b0420d87272a682bfd1e53-Paper.pdf>
- <https://arxiv.org/pdf/2007.13019.pdf>
- <https://arxiv.org/pdf/1902.10730.pdf>

# Выводы

- Важно правильно выбрать метрику в зависимости от домена и бизнес-задачи
- А/Б тест - единственный честный вариант оценить влияния рекомендаций
- Важны не сами рекомендации, а тот эффект, который они оказывают на поведение пользователя (где разместить? Новизна?)
- Лучше сразу логировать офлайн и онлайн метрики
- Увеличение офлайн метрик не всегда означает увеличение онлайн метрик
- Нужно помнить про feedback loop

# Спасибо за внимание!

Ildar Safilo

@Ildar\_Saf

irsafilo@gmail.com

<https://www.linkedin.com/in/isafilo/>